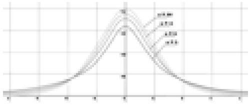
Probabilidad y Estadística

Actividades de Aprendizaje  10

Conceptos y definiciones de esta clase:

La Distribución t de Student
Introducción

Grados de Libertad
Usos

1. La Distribución t de Student

1.1 Introducción y Fórmula

La distribución t (también conocida como la t de Student) es una distribución de probabilidad que se utiliza para estimar los parámetros de la población cuando el tamaño de la muestra es pequeño y/o cuando la varianza de la población es desconocida.
Su nombre deriva del seudónimo utilizado por William S. Gosset (1876-1937) a la hora de publicar sus deducciones, puesto que por su relación contractual con la destilería Guinness estaba impedido de hacerlo, puesto que ya con anterioridad otro investigador de la firma había develado secretos industriales de la empresa y esto derivó en la prohibición de artículos por parte de los empleados.

¿Por qué utilizar la distribución t?

De acuerdo con el teorema del límite central, la distribución de muestreo de un estadístico (como la media muestral) sigue una distribución normal, siempre y cuando el tamaño de la muestra sea lo suficientemente grande. Por lo tanto, cuando conocemos la desviación estándar de la población, podemos calcular una ***medida o puntuación típica z*** (que se obtiene dividiendo la diferencia entre la medida y la media por la desviación típica del grupo), y utilizar la distribución normal para evaluar las probabilidades con la media muestral.

Sin embargo, los tamaños de muestra a veces son pequeñas, y muchas veces no sabemos la desviación estándar de la población. Cuando aparecen estos problemas, los estadísticos se basan en la distribución de la estadística t (también conocida como la puntuación t), cuyos valores son dados por:

$$t = \frac{x - \mu}{s/\sqrt{n}}$$

en donde:

- $\bar{x}$ es la media de la muestra,
- μ es la media de la población,
- s es la desviación estándar de la muestra, y

- n es el tamaño de la muestra.

La distribución de la estadística t se llama la distribución t o la distribución t de Student.

La distribución t nos permite realizar análisis estadísticos sobre ciertos conjuntos de datos que no son apropiadas para el análisis, utilizando la distribución normal.

1.2 Los grados de libertad

Existen diferentes distribuciones t. Cada forma particular de la distribución t está determinada por sus grados de libertad, que se refieren al número de observaciones independientes en un conjunto de datos.

Cuando se estima una puntuación media de una sola muestra, el número de observaciones independientes es igual al tamaño de la muestra menos uno. Por lo tanto, la distribución de la estadística t de muestras de tamaño 10 se describe mediante una distribución que tiene 10 - 1 (o 9) grados de libertad. De manera similar, en la distribución que tiene 16 grados de libertad se utiliza con una muestra de tamaño 17.

Para otras aplicaciones, los grados de libertad se pueden calcular de manera diferente.

1.3 Propiedades de la distribución t

La distribución t tiene las siguientes propiedades:

- La media de la distribución es igual a 0.
- La varianza es igual a v/ (v - 2), donde v es el grado de libertad y v> 2.
- La varianza es siempre mayor que 1, aunque muy cercana a 1 cuando hay muchos grados de libertad.
- A medida que crecen los grados de libertad, la distribución t se aproxima a la distribución normal estándar (con infinitos grados de libertad, ambas distribuciones coinciden).

1.4 ¿Cuándo utilizar la distribución t?

La distribución t se puede utilizar con cualquier estadística que tiene una distribución en forma de campana (es decir, aproximadamente normal). Las centrales de los estados límite teorema de que la distribución muestral de un estadístico será normal o casi normal, si alguna de las siguientes condiciones.

- La distribución de la población es normal.

- La distribución muestral es simétrica, unimodal, sin valores extremos, y el tamaño de la muestra es de 15 o menos.

- La distribución muestral es moderadamente asimétrica, unimodal, sin valores extremos, y el tamaño de la muestra es de entre 16 y 40.

- El tamaño de la muestra es superior a 40, sin valores atípicos.

Importante

La distribución t no debería utilizarse con muestras pequeñas en poblaciones que no son aproximadamente normales.

Ejemplo 1

Las autoridades de una determinada empresa que fabrica lámparas de bajo consumo afirman que una bombilla de luz de las que ellos fabrican dura en promedio unos 300 días. Un investigador selecciona 15 lámparas al azar para una prueba. Y determinan que las muestras duran un promedio de 290 días, con una desviación estándar de 50 días. Si la afirmación de los directivos de la empresa fuera cierta, ¿cuál es la probabilidad de que las 15 bombillas seleccionadas al azar hayan tenido una vida media de no más de 290 días?

Nota: Mostraremos dos métodos para resolver este problema, y para ambos utilizaremos una calculadora de distribución t online, como por ejemplo la que aparece en el siguiente vínculo, aunque el lector puede utilizar la que desee.

Links a temas de interés

- http://www.danielsoper.com/statcalc3/calc.aspx?id=41
- http://www.usablestats.com/calcs/tdist
- http://www.usablestats.com/calcs/1samplet&summary=1

Primera solución (enfoque tradicional)

Calcularemos la puntuación t, sobre la base de datos que se presentan en la descripción del problema, y a continuación, utilizaremos la calculadora de distribución T para encontrar la probabilidad.

Puntuación t

$$t = \frac{x - \mu}{s/\sqrt{n}}$$

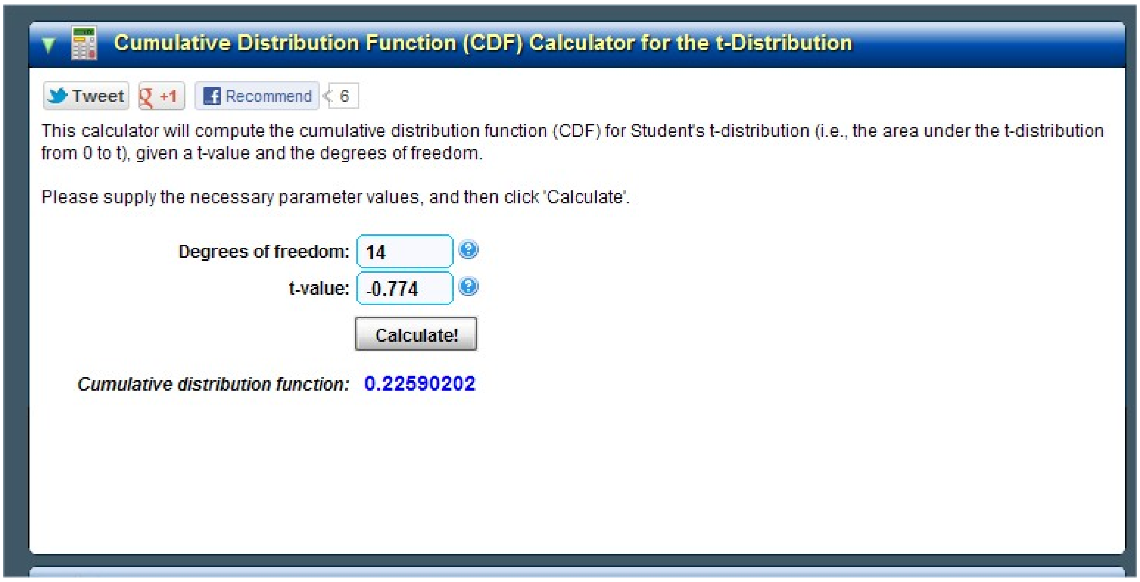
$$t = \frac{290 - 300}{50/\sqrt{15}} = \frac{-10}{12.9099} = -0.7745966$$

Ahora, estamos listos para utilizar la calculadora de la distribución t.

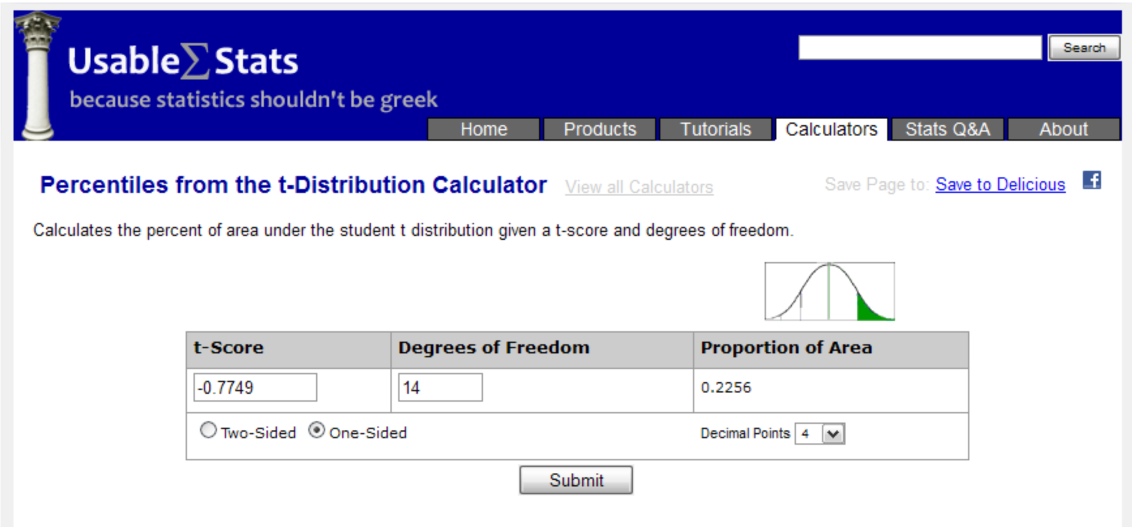
Los grados de libertad son iguales a 15 - 1 = 14

La puntuación T es igual a - 0,7745966.

A continuación, mostraremos cómo fueron ingresados estos datos en la página y cuál es el resultado obtenido:



Si desea utilizar el segundo enlace proporcionado los resultados se verían así:



En ambos casos la calculadora muestra la probabilidad acumulada de aproximadamente 0.226. Por lo tanto, si es cierto que la vida de la bombilla es de 300 días, hay una probabilidad del 22,6% que la vida media de 15 bulbos seleccionados al azar sea menor que o igual a 290 días.

Segunda solución

En esta solución sólo tiene que introducir los datos del problema en la calculadora de la distribución t

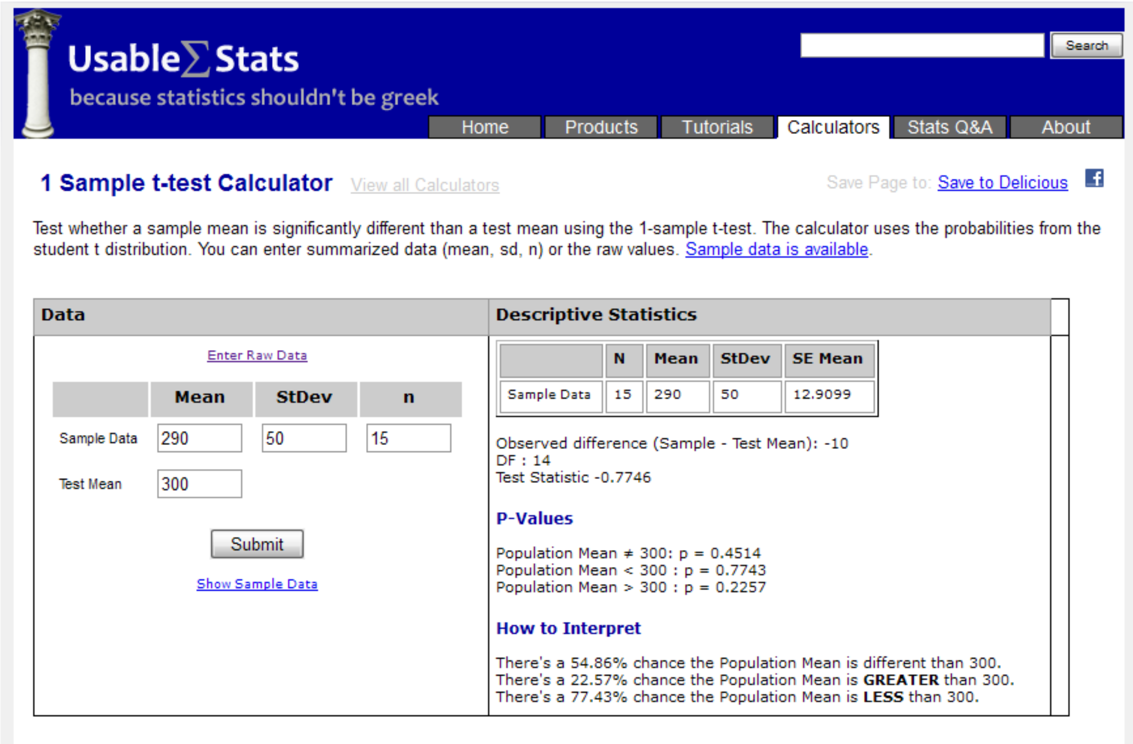
Grados de libertad: 15 - 1 = 14

Media de la población: 300

Media de la muestra: 290

Desviación estándar de la muestra: 50

Para este caso, utilizaremos el tercer link proporcionado, en el cual una vez que usted ingrese los datos del problema, se verá así:



En este caso en particular, la calculadora no sólo nos devuelve los datos solicitados, sino que además nos muestra la interpretación de los mismos, donde puede leerse que existe una probabilidad del 22,6% de que en promedio un bulbo de la muestra extraída se queme dentro de 290 días.

Importante

Observe bien los datos, identificando cada uno; anote los resultados y compare con sus interpretaciones.

A continuación, le planteamos un problema similar para que pueda resolverlo y lo compare con las respuestas que le ofrecemos.

Situación Problemática 1

Suponga que las puntuaciones en un test de inteligencia se distribuyen normalmente, con una media de 100. Supongamos que se seleccionan al azar 20 personas y se les toma el test. La desviación estándar en el grupo es de 15. ¿Cuál es la probabilidad de que la puntuación promedio de la prueba en el grupo sea como máximo 110?

Solución:

Lo invitamos a que detecte los datos del problema y los ingrese en una calculadora online. Si tiene dificultad para identificar los datos en el problema, siga leyendo.

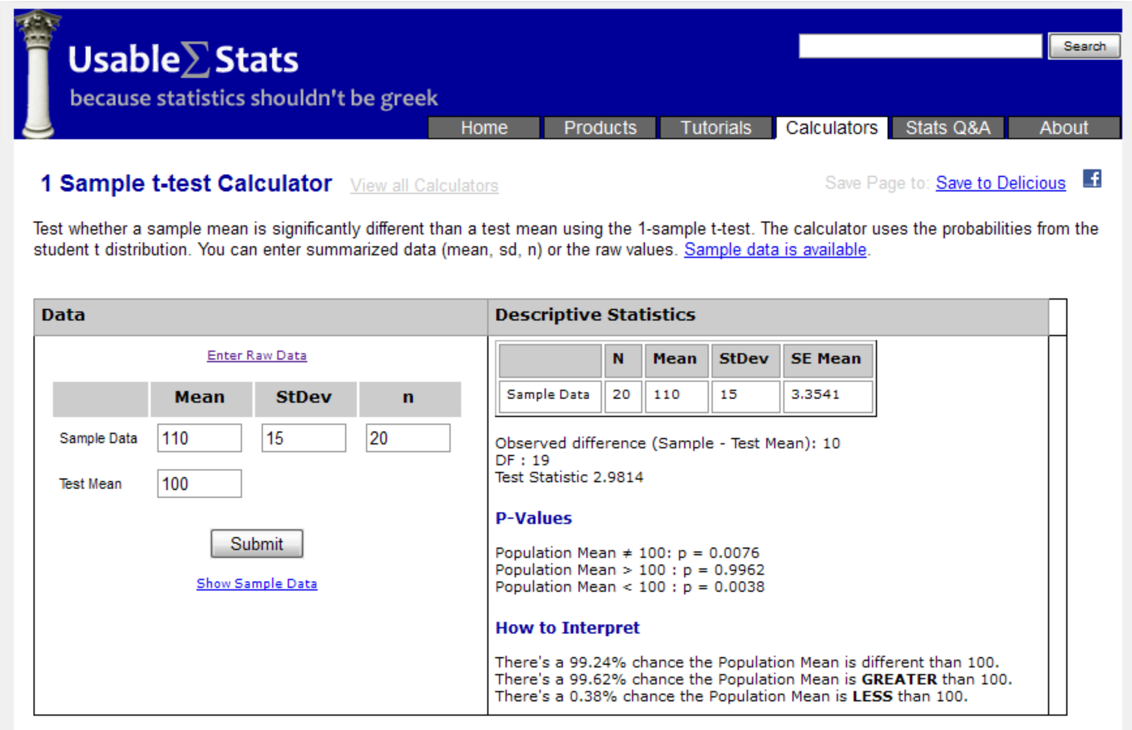
Los grados de libertad son iguales a 20 - 1 = 19.

La media de la población es igual a 100.

La media de la muestra es igual a 110.

La desviación estándar de la muestra es de 15.

Ingresamos estos valores en la calculadora de la distribución t. La calculadora muestra la probabilidad acumulada: 0.996. Por lo tanto, existe una probabilidad del 99,6% que el promedio de la muestra no será mayor que 110.



Situación Problemática 2

Determinar los valores de una variable t con 16 grados de libertad:

a) que deja a su derecha un área 0'01

b) que deja a su izquierda un área 0'05

c) Determinar el valor de una variable t con 90 grados de libertad que deja a su derecha un área igual a 0'121.

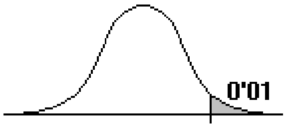
Calcule, con 20 grados de libertad:

d) Pr (|t| < 2'09)

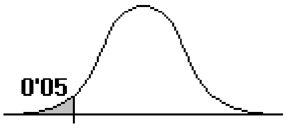
e) Pr (1'3 ≤ t ≤ 2'5)

Soluciones:

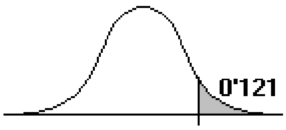
a) De la tabla obtenemos directamente el valor t = 2'583476



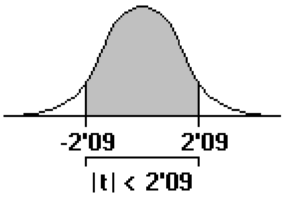
b) Por la simetría de la distribución, buscamos el valor t que deja a su derecha un área 0'05. El valor pedido es el opuesto del anterior: t = -1'745880



c) Aproximamos mediante la distribución normal N (0,1). Entre 0 y el valor pedido queda un área 0'5 - 0'121 = 0'379, correspondiendo al valor z = 1'17. Luego: t ≈ z = 1'17



d) El valor más próximo de t es 2'085962. De aquí, consultando la tabla que relaciona t con las dos áreas laterales:
P (t < 2'05) ≈ 1 - 0'05 = 0'95



e) Los valores más próximos a los de t (con 20 g. l.) son, respectivamente, 1'325340 y 2'527973, con áreas a su izquierda 0'90 y 0'99.
Luego: P (1'3 ≤ t ≤ 2'5) ≈ 0'99 - 0'90 = 0'09